Sections 10.1-10.3

Chester Ismay, Tom Linton

Ripon College, Central College

"The brain is like a muscle. When it is in use we feel very good. Understanding is joyous."

- Carl Sagan

(1) For used cars that are for sale, if one studies the association between the mileage of the car (as indicated by the odometer) and asking price, what direction would you expect the association to have?

A Negative

B No Association

C Positive

D You cannot tell

E I am unsure

(2) For new employees, if a company makes a scatterplot of $x =$ hours spent training and $y =$ number of mistakes made the first day after training, what direction would you expect the association to have?

A Negative

B No Association

C Positive

D You cannot tell

E I am unsure

(3) Which of the following pairs of variables would you expect to have the weakest association (correlation near zero)?

A For baseball players, number of at bats and number of hits.

B For colleges, the winning percentage of the volleyball team and the number of fulltime students.

C Length of a race and average velocity of runners in the race.

D For cities, distance to Milwaukee and airfare from Milwaukee to the city.

E All of the above would have fairly strong associations.

(4) If the Ripon College golf team played a two–day tournament where each golfer scored 2 strokes lower on day 2 then they did on day 1, what would be the value of the correlation coefficient between $x =$ score on day 1 and $y =$ score on day 2?

A $r = -2$, since day 2 scores are lower by 2 strokes.

B $r = -1$, since this describes a line with negative slope.

C $-1 \leq r \leq 1$, this is all one can say here.

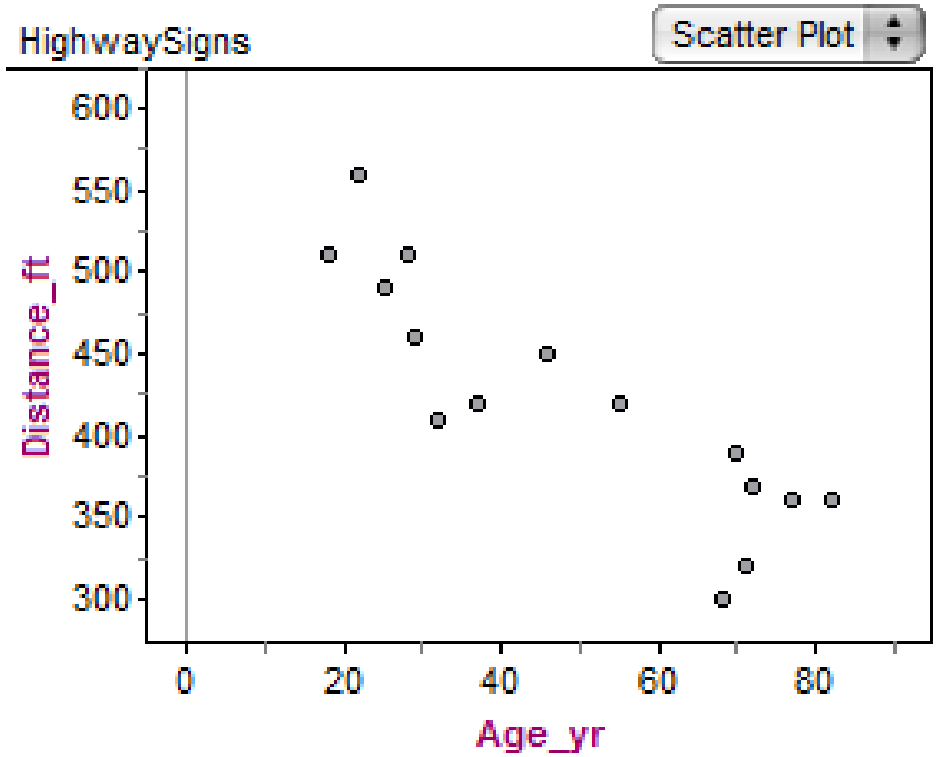
D $r = 1$, since this describes a line with positive slope.

E $r > 0$, this describes a positive association but the strength is unclear.

Setup

In a study of legibility and visibility of highway signs, a Pennsylvania research firm determined the maximum distance (in feet) at which each of 15 drivers could read a newly designed highway sign. The firm also recorded the age (in years) of each of the drivers. The 15 drivers in the study ranged from 18 to 82 years old. The firm wanted to investigate if there was an association between a driver's age and the maximum distance from which they could read the new style of sign. They believed that younger drivers could read the sign from a farther distance than older drivers.

A scatterplot of the data is shown here.



(5) What is the explanatory variable?

A The age of the drivers.

B The maximum distance of readability.

C Legibility and visibility of highway signs.

D If reading distance decreases with age.

E The average age of the drivers.

(6) A scatterplot of the data is shown below. Which of the following could be the value of the correlation coefficient for this study?

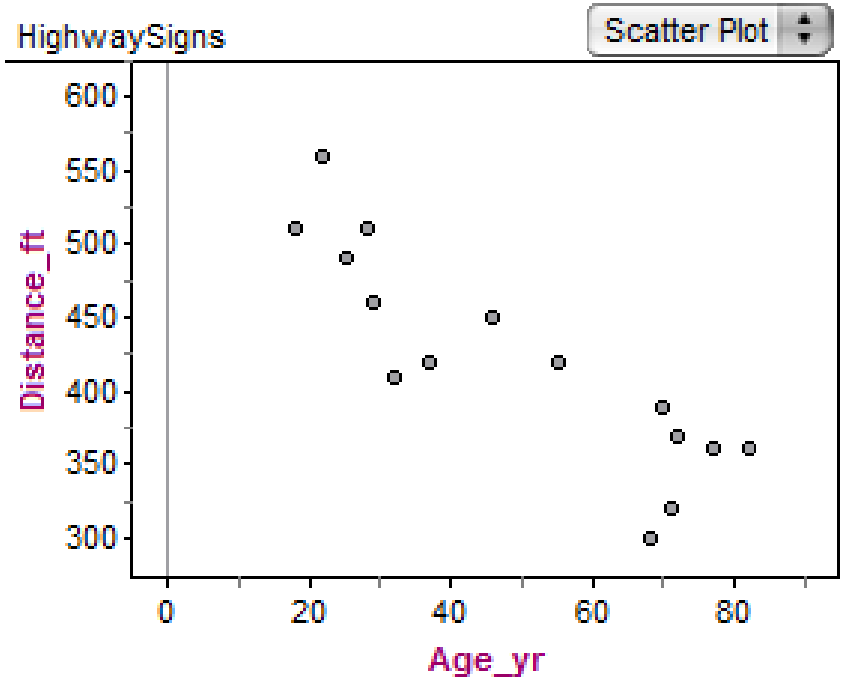
A $r = 0.9$

B $r = 0.3$

C $r = 0$

D $r = -0.3$

E $r = -0.9$



(7) Describe the strength of the linear association between age and reading distance.

A Strong

B Moderate

C Weak

D Negative

E Highly curved

(8) The least squares regression equation for predicting reading distance (in feet) from age (in years) is $\widehat{distance} = 561.40 - 2.86\, age$. Which is the best (most complete) interpretation of the slope of this line?

A As age increases, the maximal reading distance drops.

B A newborn could read the sign from about 561 feet away.

C For each additional year in age, the predicted maximal reading distance decreases by 2.86 feet on average.

D As the age of drivers increases from 18 to 82 years old, the reading distance drops by 286%.

E Each increase of one year in age is associated with an average predicted decrease in reading distance of 2.86%.

Based on data from early in the 2004 PGA season, the least squares regression equation for predicting a professional golfer's scoring average (in strokes per round, recall that low scores are better in golf), based on their average driving distance (in yards) is $\widehat{score} = 77.79 - 0.025\, drivingDistance$, with $r = 0.2655$ and $r^2 = 0.07$. A simulated significance test to help decide if the population correlation is zero or not, gave $p = 0.017$.

(9) Based on the summary statistics given, what does this small p–value ($p = 0.017$) reveal?

A It reveals evidence of a strong association.

B It reveals strong evidence of an association.

C It reveals strong evidence of a strong association.

D It reveals weak evidence of an association.

E None of the above.

(10) Based on the summary statistics given, do these data suggest that one can significantly improve the prediction of a golfer's scoring average by using their average driving distance, as opposed to just using the overall scoring average for all PGA golfers?

A Yes, the p–value is small, so the results are significant.

B Yes, the slope is negative, so longer drives are associated with lower scoring averages.

C No, the p–value is small, so a correlation of zero is plausible.

D Not much will be gained since $r^2 = 0.07$.

E You need more information to decide.

(11) What is the best interpretation of the $r^2 = 0.07$ value in this problem?

A The correlation between scoring average and driving distance.

B An increase of one unit in driving distance corresponds to r^2 units in scoring average

C The percentage of total variation in the scoring average that is explained by the linear relationship with driving distance.

D Two of the above are correct.

E None of the above.

Key Terms and Ideas to Understand in Chapter 10

- Correlation coefficient
- Extrapolation
- Influential observations
- r versus ρ
- b versus β
- Residuals
- Slope
- Tactile experiment for simulation
- y-intercept